AD-A031 477 CASE WESTERN RESERVE UNIV CLEVELAND OHIO DEPT OF MA--ETC F/G 5/9
ADAPTIVE FORECASTING OF THE SIZE OF A FORCE SUBJECT TO RANDOM W--ETC(U)
OCT 76 S ZACKS, S E HABER N00014-75-C-0529

UNCLASSIFIED TR-27 NL

1 OF 1
ADA031477

END
DATE
FILMED
12 - 76

FG

12

ADAPTIVE FORECASTING OF THE SIZE OF
A FORCE SUBJECT TO RANDOM WITHDRAWALS

by

S. Zacks and S. E. Haber

1473 in pouch

Technical Report No. 27

October 15, 1976

PREPARED UNDER CONTRACT
N00014-75-C-0529, PROJECT NR 042-276
OFFICE OF NAVAL RESEARCH

Department of Mathematics and Statistics
CASE WESTERN RESERVE UNIVERSITY
Cleveland, Ohio

ADAPTIVE FORECASTING OF THE SIZE OF A FORCE SUBJECT TO RANDOM WITHDRAWALS[1]

by

S. Zacks and S. E. Haber

Case Western Reserve University

and

The George Washington University

1. Introduction

In a recent technical paper [4] a method of forecasting the size of force (military organization) subject to random withdrawals was provided. The method presented there considers the records on the size of the force in six months intervals, estimates the retention rates of certain subgroups (cohorts) at the time of forecasting and provides a prediction interval on the anticipated size of the force six or twelve months ahead. The method developed in [4] was specifically oriented to the problem of forecasting the size of the Marine Corps and applied in [5] and [6] to estimate the total size of the Marine Corps. As explained in [4], we distinguish between four phases in the service of an enlistee. Phase I consists of the first six months of basic training; Phase II follows Phase I and continues until six months before the termination of the (first)

[1] Supported by the Office of Naval Research under Contract N00014-75-C-0728, Project NR 347-020 at the Program in Logistics, The George Washington University and under Contract N00014-75-C-0529, Project NR 042-276 at Case Western Reserve University

contract. Phase III consists of the last six months of the (first) contract. Phase IV is the phase of career service. We have shown in [4] that the first three phases are similar to those of reliability systems, in which the failure rate function attains after an initial phase of high failure rate a phase of stable failure rate (Phase II) and then reaches a phase of wearout (Phase III). In the present paper we develop the theory for Phase II forecasting and compare three alternative procedures. One procedure provides conditional maximum likelihood estimates (CMLE) of the limits of prediction intervals. The second procedure determines tolerance intervals and the third procedure provides Bayes prediction intervals for the size of the force. We show that although the CMLE-procedure is less conservative than the other two procedures, it has yielded prediction intervals for the total force which contain the actual values (on retroactive basis). This is due to the fact that the samples are large and the standard errors of the estimates are sufficiently small.

The Phase II forecasting is performed for cohorts consisting of enlistees classified according to the following factors: time of entry, length of contract, race and education. Haber has shown in [2] that among many factors that can possibly affect the retention rates the above factors are most significiant. We base the model of the present study on the same factors. The structure of the data as related to these factors is explained in Section 2. In Section 3 we discuss the basis statistical model for adaptive forecasting. In Section 4 we develop the concept of prediction intervals as a tool of forecasting. In Section 5 we discuss CMLE-prediction limits, tolerance limits and Bayes prediction limits for the size of cohorts which have been in Phase II more than six

months (at least 12 months in service). All the procedures of forecasting are compared numerically by applying them retroactively on actual Marine Corps data for the period January 1969 to July 1972. Section 6 is devoted to the forecasting problem for cohorts just entering Phase II. The methods developed are related to those of Section 5. FORTRAN programs according to which the examples have been computed are available and can be obtained upon request. Finally we mention that although there is vast literature on subjects which are touched on or mentioned in the present study, either from the point of view of man-power statistics or from the areas of reliability, Poisson processes, statistical distributions and Bayesian analysis we have cited in the paper specifically only references which are most relevant to the discussion.

2. The Data and Structural Components

The random variables considered are the number of enlistees remaining in service at six months intervals, in group (cohorts) classified according to the following characteristics:

A. Length of enlistment contract (LC)

Two years,	$i = 1$
Three years,	$i = 2$
Four years,	$i = 3$

B. Education Level (ED)

Less than high school,	$j = 1$ (LHS)
High school or more,	$j = 2$ (AHS)

C. Race (R)

White,	$k = 1$ (W)
Non-White,	$k = 2$ (NW)

D. Time of Entry into Service (in six months periods) (EN)

January 1-June 30, 1969	$\ell = 1$	(1-69)
July 1-December 31, 1969	$\ell = 2$	(7-69)
Up to January 1-June 30, 1972	$\ell = 7$	(1-72)

In the following table we present the observed values of these random variables at the time intervals starting on July 1, 1971, January 1, 1972 and July 1, 1972. These values are later used to illustrate the forecasting procedures.

As in the previous study [4] we distinguish between three phases in the service of first enlistees.

Phase I: the first six months of basic training.

Phase II: the period following Phase I up the the last six months before termination of contract.

Phase III: the last six months of the contract.

In Figure 1 we illustrate the retention of cohorts (in percents) in relation to their three phases of service.

Figure 1.

Retention Of First Enlistees (in Percents) Among Four-year, Non-White Above High School Cohorts

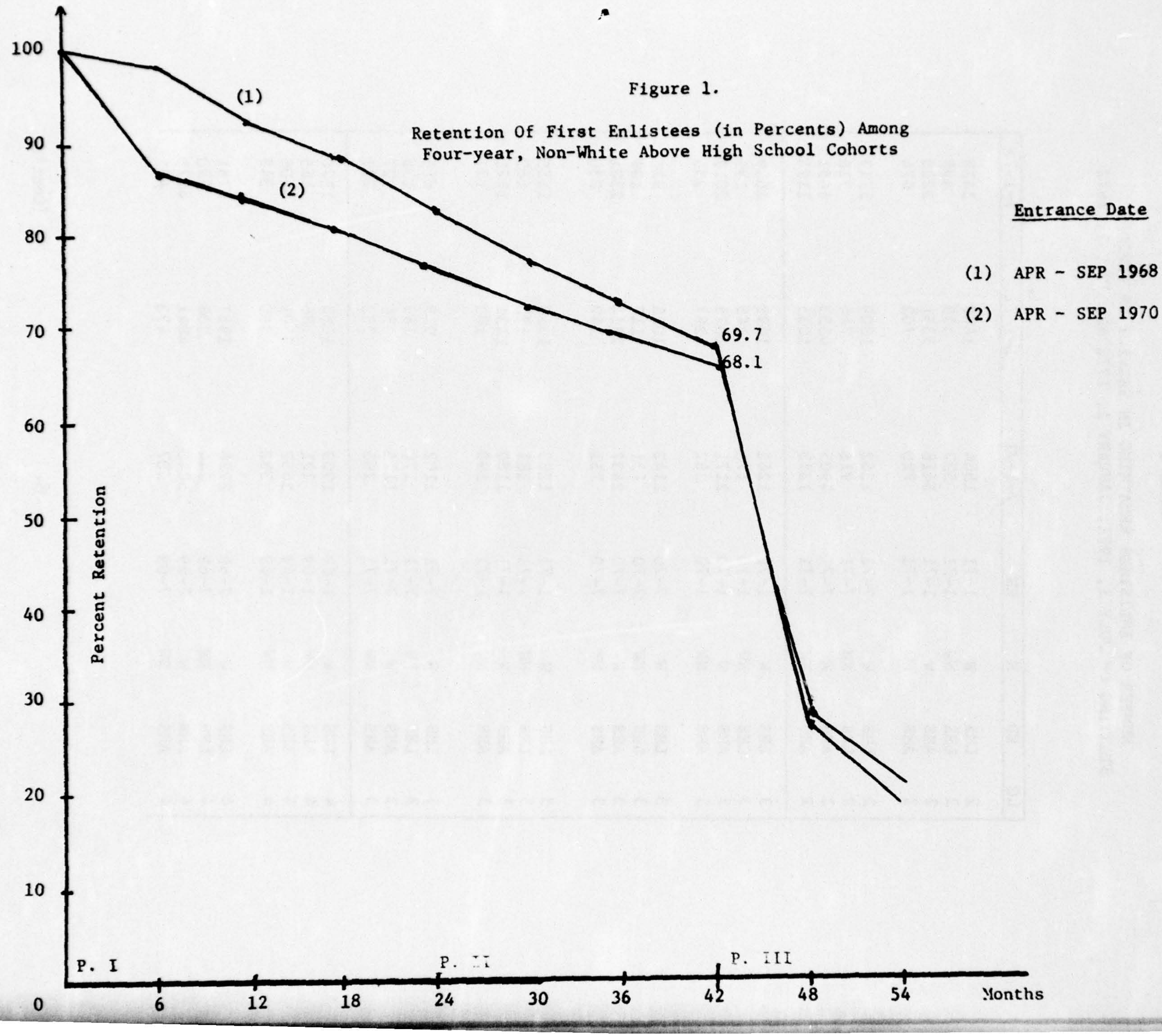

TABLE 1

NUMBER OF ENLISTEES REMAINING IN SERVICE IN PERIODS STARTING ON JULY 1, 1971, JANUARY 1, 1972 AND JULY 1, 1972

LC	ED	R	EN	7-1-71	1-1-72	7-1-72
2	LHS	W	1-71	1564	1489	1418
2	LHS	NW	1-71	537	525	499
2	AHS	W	1-71	3416	3351	3280
2	AHS	NW	1-71	720	702	674
2	LHS	W	7-71	2362	1809	1717
2	LHS	NW	7-71	916	744	716
2	AHS	W	7-71	4905	4553	4451
2	AHS	NW	7-71	1353	1235	1195
3	LHS	W	1-70	1262	1099	1009
3	LHS	NW	1-70	173	145	135
3	AHS	W	1-70	2171	2079	2019
3	AHS	NW	1-70	282	267	256
3	LHS	W	7-70	1143	1012	910
3	LHS	NW	7-70	131	117	104
3	AHS	W	7-70	2437	2364	2301
3	AHS	NW	7-70	251	240	230
3	LHS	W	1-71	1293	1219	1155
3	LHS	NW	1-71	181	172	163
3	AHS	W	1-71	1180	1151	1121
3	AHS	NW	1-71	190	183	177
3	LHS	W	7-71	1102	919	866
3	LHS	NW	7-71	170	143	133
3	AHS	W	7-71	1115	1034	997
3	AHS	NW	7-71	260	233	227
4	LHS	W	1-69	1909	1688	1570
4	LHS	NW	1-69	221	186	165
4	AHS	W	1-69	3652	3501	3408
4	AHS	NW	1-69	382	360	345
4	LHS	W	7-69	2226	1957	1781
4	LHS	NW	7-69	244	208	190
4	AHS	W	7-69	5040	4841	4675
4	AHS	NW	7-69	507	473	451

(Cont'd)

TABLE 1 (Cont'd)

LC	ED	R	EN	7-1-71	1-1-72	7-1-72
4	LHS	W	1-70	2935	2594	2347
4	LHS	NW	1-70	384	327	292
4	AHS	W	1-70	3715	3556	3416
4	AHS	NW	1-70	489	454	428
4	LHS	W	7-70	3173	2864	2633
4	LHS	NW	7-70	356	321	294
4	AHS	W	7-70	4401	4267	4108
4	AHS	NW	7-70	547	511	478
4	LHS	W	1-71	3325	3146	2979
4	LHS	NW	1-71	423	410	386
4	AHS	W	1-71	3366	3294	3207
4	AHS	NW	1-71	559	546	526
4	LHS	W	7-71	3969	3358	3174
4	LHS	NW	7-71	479	402	383
4	AHS	W	7-71	4600	4445	4281
4	AHS	NW	7-71	688	632	607

3. The Statistical Model

Let $X_{ijk\ell}(t)$ denote the number of enlistees remaining in service at the end of the t-th epoch, where an epoch is a six-month period, from the group (cohort) combined of the (i,j,k,ℓ) factors-combination. The statistical model specifies that the conditional distributions of $X_{ijk\ell}(t+1)$ given $X_{ijk\ell}(t)$ is the binomial distribution with parameters $X_{ijk\ell}(t)$ (number of Bernoulli trails) and $\theta_{ijk\ell}(t+1)$ (retention probability) i.e., for every $t \geq 1$, and all (i,j,k,ℓ)

$$X_{ijk\ell}(t+1)|X_{ijk\ell}(t) \sim B(X_{ijk\ell}(t),\theta_{ijk\ell}(t+1)). \tag{3.1}$$

Generally $B(N,p)$ denotes a bionomial distribution corresponding to N Bernoulli trials and probability of success p, $0 < p < 1$. The symbol $\sim$ in (3.1) designates that, conditionally on $X_{ijk\ell}(t)$, the random

variable $X_{ijk\ell}(t+1)$ has the specified binomial distribution. Indeed, the number of enlistees from the (i,j,k,ℓ) cohort remaining in service at the (t+1)st epoch can assume any value from 0 to $X_{ijk\ell}(t)$. It is assumed that each individual in a given cohort has the same retention probability in a given future period and that the events of retention or withdrawal of different individuals are independent. This model may be justifiably criticized if applied to small groups, due to possible interactive forces among the members. It seems, however, to be an adequate model for describing the retention in large cohorts.

4. Prediction Intervals

After observing the sequence of X-values for each (i,j,k,ℓ) combination, at epochs $1,\ldots,t$, we wish to forecast their possible values s epochs in the future. One can easily verify that the conditional distribution of $X_{ijk\ell}(t+s)$, given $X_{ijk\ell}(t)$ is the binomial $B(X_{ijk\ell}(t),\ P_{ijk\ell}(t,s))$ where,

$$P_{ijk\ell}(t,s) = \prod_{v=1}^{s} \theta_{ijk\ell}(t+v),\ t = 0,1,\ldots,\ s = 1,2,\ldots\ . \tag{4.1}$$

Accordingly, if the sequence of retention probabilities $\underset{\sim}{\theta}_{ijk\ell} = \{\theta_{ijk\ell}(t);\ t \geq 1\}$ is known one can determine for each t and s prediction intervals $(\underline{\Pi}_{ijk\ell}(t+s), \overline{\Pi}_{ijk\ell}(t+s))$, satisfying the requirement that, for each $t \geq 0$

$$P[\underline{\Pi}_{ijk\ell}(t+s) \leq X_{ijk\ell}(t+s) \leq \overline{\Pi}_{ijk\ell}(t+s) \mid X_{ijk\ell}(t), \underset{\sim}{\theta}_{ijk\ell}\} \geq 1-\alpha \tag{4.2}$$

$(1 - \alpha)$ is a preassigned probability level.

The limits of these prediction intervals depend on $X_{ijk\ell}(t)$ and

on $P_{ijk\ell}(t,s)$ and are given by $\underline{\eta}_{ijk\ell}(t+s) =$

$$\underline{\eta}_{ijk\ell}(t+s) = \underline{\text{maximal non-negative integer}},\ k, \quad \underline{\text{such that}}\ \ B(k|X_{ijk\ell}(t), P_{ijk\ell}(t,s)) \le \frac{\alpha}{2} \tag{4.3}$$

and

$$\bar{\eta}_{ijk\ell}(t+s) = \underline{\text{least non-negative integer}},\ k, \quad \underline{\text{such that}}\ \ B(k|X_{ijk\ell}(t), P_{ijk\ell}(t,s)) \ge 1 - \frac{\alpha}{2}\ ; \tag{4.4}$$

where $B(k|X_{ijk\ell}(t), P_{ijk\ell}(t,s))$ is the c.d.f. of the corresponding binomial distribution. The determination of prediction limits (4.3) and (4.4) requires the computation of the corresponding binomial c.d.f.'s. In the cases considered here, the X-values are generally in the order of several hundreds or thousands (see Table 1). In these cases the numerical determination of the binomial c.d.f. applies the normal approximation (see Johnson and Kotz [3; pp. 62]). Based on this approximation we employ the following large sample formulae for the prediction limits

$$\bar{\eta}_{ijk\ell}(t+s) = X_{ijk\ell}(t)P_{ijk\ell}(t,s) + z_{1-\alpha/2}\left\{X_{ijk\ell}(t)\cdot P_{ijk\ell}(t,s)\cdot Q_{ijk\ell}(t,s)\right\}^{1/2}$$

and (4.5)

$$\underline{\eta}_{ijk\ell}(t+s) = X_{ijk\ell}(t)P_{ijk\ell}(t,s) - z_{1-\alpha/2}\left\{X_{ijk\ell}(t)P_{ijk\ell}(t,s)Q_{ijk\ell}(t,s)\right\}^{1/2}$$

where $Q_{ijk\ell}(t,s) = 1 - P_{ijk\ell}(t,s)$ and $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$th fractile of the standard normal distribution. Generally the values of the retention probabilies $P_{ijk\ell}(t,s)$ are unknown and should be estimated from the data. In the following sections we discuss the problem of determining the prediction limits when the retention probabilities are unknown.

5. Estimating the Prediction Intervals

In the present section we duscuss three methods of estimating the prediction intervals, when the retention probabilities are unknown. For the sake of simplifying the presentations we restrict attention to six months forecasting, $s = 1$. We discuss here only cohorts in Phase II, which have been in service at least 12 months, i.e., $t \geq 2$. The case $t = 1$ will be discussed later.

5.1 CMLE-Prediction Limits

For estimating the retention probabilities $\theta_{ijk\ell}(t+1)$ $t \geq 2$, we develop methods which depend only on the observed statistics of the cohort under consideration. These methods can be modified to apply also statistics of previous cohorts (as will be indicated in Section 6). For this purpose we have to introduce the assumption that all the cohorts having the same length of contract and the same education race combination behave similarly, with respect to their attrition, irrespective of the period of entry to service. The methods discussed in the present section are free of such an assumption. We have to assume, however, that the retention probabilities of each cohort have close values in adjacent periods. Thus, we can estimate $\theta_{ijk\ell}(t)$ by the conditional maximum likelihood (CMLE) estimates

$$\hat{\theta}_{ijk\ell}(t) = X_{ijk\ell}(t)/X_{ijk\ell}(t-1), \ t \geq 2. \tag{5.1}$$

This estimator employes, for each cohort, only the current and the previous six months values of X. Point estimates of the prediction limits for $X_{ijk\ell}(t+1)$ are obtained by substituting in formula (4.3) the values of and

$\hat{\theta}_{ijk\ell}(t)$ for $P_{ijk\ell}(t,s)$. These estimates of the prediction limits are labelled the CMLE-prediction limits. They are given by the formula

$$X_{ijk\ell}(t) \cdot \hat{\theta}_{ijk\ell}(t) \pm \tag{5.2}$$

$$z_{1-\alpha/2} \left\{ X_{ijk\ell}(t)\hat{\theta}_{ijk\ell}(t) \cdot (1 - \hat{\theta}_{ijk\ell}(t)) \right\}^{1/2} .$$

Indeed, according to the invariance principle of maximum likelihood estimation (see Zacks [7; pp. 223]) the CMLE-limits given by (5.2) are themselves conditional maximum likelihood estimators of the prediction limits. Furthermore, according to the large sample theory of maximum likelihood estimators (see Zacks [7; pp. 247]) the sampling distribution of these CMLE limits are approximately normal around the true prediction limits. Accordingly we may expect that in about 25-50% of the cases the CMLE-prediction intervals (5.2) will be too short and will not cover the true prediction intervals. This is our main reason for developing alternative methods of estimation. On the other hand, due to the large samples under consideration and the strong consistency of the maximum likelihood estimators (see Zacks [7]) we expect that the CMLE-prediction limits will be close to the true limits, even if the intervals are not entirely covered. Numerical examples will be given in the sequel to illustrate this point.

5.2 Tolerance Limits for the Prediction Intervals

One type of tolerance limits can be defined as confidence limits to fractiles of distribution. We derive here tolerance limits for the prediction intervals (5.2) within a large sample framework. It is well known that in large binomial samples, if $\hat{\theta}$ is the CMLE of θ then

$2 \sin^{-1}(\sqrt{\hat{\theta}})$ has approximately normal distribution with expectation $2 \sin^{-1}(\sqrt{\theta})$ and variance $1/n$, where n is the sample size (see Johnson and Kotz [3; pp. 65]). Since the conditional distribution of $X_{ijk\ell}(t+1)$ given $X_{ijk\ell}(t)$ is binomial with retention probability $\theta_{ijk\ell}(t)$ we consider the statistic

$$Y_{ijk\ell}(t) = 2 \sin^{-1}\left(\sqrt{\frac{X_{ijk\ell}(t)+.5}{X_{ijk\ell}(t-1)+1}}\right), \ t \geq 2 \tag{5.3}$$

The mean of the asymptotic conditional normal distribution of $Y_{ijk\ell}(t)$, given $X_{ijk\ell}(t-1)$, is

$$H_{ijk\ell}(t) = 2 \sin^{-1}\left(\sqrt{\theta_{ijk\ell}(t)}\right), \tag{5.4}$$

and its asymptotic conditional variance is $1/X_{ijk\ell}(t-1)$.

Let $Y_{ijk\ell}(t+1)$ represent a statistic as in (5.3) based on $X_{ijk\ell}(t)$ and the yet unobserved $X_{ijk\ell}(t+1)$. The asymptotic conditional distribution of $Y_{ijk\ell}(t+1)$ given $X_{ijk\ell}(t)$ is, according to our assumption that $\theta_{ijk\ell}(t+1) = \theta_{ijk\ell}(t)$, normal with mean $H_{ijk\ell}(t)$ and variance $1/X_{ijk\ell}(t)$.

We determine now confidence limits for the unknown value of $H_{ijk\ell}(t)$. A $(1 - \alpha/2)$ lower confidence limit for $H_{ijk\ell}(t)$ is, according to the normal approximation,

$$L_{ijk\ell}(t) = Y_{ijk\ell}(t) - z_{1-\alpha/2} / \sqrt{X_{ijk\ell}(t-1)}. \tag{5.5}$$

Similarly, a $(1 - \alpha/2)$, upper confidence limit for $H_{ijk\ell}(t)$ is approximately

$$U_{ijk\ell}(t) = Y_{ijk\ell} + z_{1-\alpha/2} / \sqrt{X_{ijk\ell}(t-1)}. \tag{5.6}$$

We apply now the Bonferoni inequality

$$P[A \cap B] \geq 1 - P(\bar{A}) - P(\bar{B}), \tag{5.7}$$

where A is the event that the confidence limits (5.4) and (5.5) cover $H_{ijk\ell}(t)$ and B is the event that $Y_{ijk\ell}(t+1)$ belongs to the interval

$$H_{ijk\ell}(t) \pm z_{1-\alpha/2} \Big/ \sqrt{X_{ijk\ell}(t)} \tag{5.8}$$

Accordingly, the probability is at least $1 - 2\alpha$ at the limits

$$Y_{ijk\ell}(t) \pm z_{1-\alpha/2}\left(\frac{1}{\sqrt{X_{ijk\ell}(t-1)}} + \frac{1}{\sqrt{X_{ijk\ell}(t)}}\right) \tag{5.9}$$

cover the future value $Y_{ijk\ell}(t+1)$. The limits will be called <u>tolerance limits</u>. Finally, since the function $2 \sin^{-1}(\sqrt{\theta})$, for $0 < \theta < 1$, is strictly increasing we employ the inverse of this transformation on the two tolerance limits of $Y_{ijk\ell}(t+1)$ to obtain corresponding tolerance limits for $X_{ijk\ell}(t+1)$. These limits are:

$$\xi^{(U)}_{ijk\ell}(t) = (\sin(Y^{(U)}_{ijk\ell}/2))^2 \cdot (X_{ijk\ell}(t) + 1.) - .5 , \tag{5.10}$$

and

$$\xi^{(L)}_{ijk\ell}(t) = (\sin(Y^{(L)}_{ijk\ell}/2.))^2 \, (X_{ijk\ell}(t) + 1.) - .5 , \tag{5.11}$$

where $Y^{(U)}_{ijk\ell}(t)$ and $Y^{(L)}_{ijk\ell}(t)$ are the upper and lower tolerance limits of $Y_{ijk\ell}(t+1)$.

5.3 <u>Bayes Prediction Limits</u>

The forecasting of $X_{ijk\ell}(t+1)$, given the prior data, can be attempted also from a Bayesian point of view. The limits of the prediction

intervals, as defined in Section 4, depend on the unknown retention probabilities $\theta_{ijk\ell}(t)$. In a Bayesian framework we ascribe these parameters prior distributions. Given the past observations we can derive the posterior distributions of $\theta_{ijk\ell}(t)$. Let $G^{(t)}_{ijk\ell}(\theta)$ denote the prior distribution function of $\theta_{ijk\ell}(t)$, $t = 2,3,\ldots$. We assume that for each (i,j,k,ℓ) the retention probabilities $\theta_{ijk\ell}(t)$ and $\theta_{ijk\ell}(t+1)$ have the same prior distributions. Let $G^{(t)}_{ijk\ell}(\theta|X)$ denote the posterior distribution of $\theta_{ijk\ell}(t)$ given that $X_{ijk\ell}(t) = X$. According to this posterior distribution we can compute the posterior marginal distribution of $X_{ijk\ell}(t+1)$ given $X_{ijk\ell}(t) = X$. This distribution is called also the predictive distribution and is given by

$$\Pi^{(t)}_{ijk\ell}(\xi|X) = \int P\{X_{ijk\ell}(t+1) \le \zeta|X_{ijk\ell}(t) = X, \theta_{ijk\ell}(t+1) = \theta]dG^{(t)}_{ijk\ell}(\theta|X). \tag{5.12}$$

Bayes prediction limits for $X_{ijk\ell}(t+1)$, at level $1 - \alpha$, are values $\underline{\zeta}$ and $\bar{\zeta}$ such that

$$\Pi^{(t)}_{ijk\ell}(\zeta|X) - \Pi^{(t)}_{ijk\ell}(\underline{\zeta}|X) \ge 1 - \alpha. \tag{5.13}$$

Aitchison and Dunsmore [1] consider for binomial distributions of future observations, beta-binomial predictive distributions, given the number of successes in the past (see [1] pp. 24). These are predictive distributions obtained by ascribing θ prior beta distributions. In our specific problem it would be impractical to consider such beta-binomial predictive distributions from the numerical point of view. Instead we apply the arcsin transformations (5.3). As before, the conditional (large sample) distribution of $Y_{ijk\ell}(t+1)$ given $X_{ijk\ell}(t)$ is normal with expectation

$H_{ijk\ell}(t) = 2\sin^{-1}(\sqrt{\theta_{ijk\ell}(t)})$, and variance $1/X_{ijk\ell}(t)$. We assume that the prior distribution of $H_{ijk\ell}(t)$ is normal with expectation w and variance τ^2, i.e., $H_{ijk\ell}(t) \sim N(w,\tau^2)$. It is immediate then to verify that the posterior distribution of $H_{ijk\ell}(t)$, given that $Y_{ijk\ell}(t) = y$ and $X_{ijk\ell}(t-1) = x$ is

$$H_{ijk\ell}(t)\,|\,Y_{ijk\ell}(t) = y, X_{ijk\ell}(t-1) = x \tag{5.14}$$

$$\sim N\left(\frac{y\tau^2 x + w}{1 + \tau^2 x}, \frac{\tau^2}{1 + \tau^2 x}\right)$$

The symbol $\sim$ designates "distributed like". (5.14) means that the posterior distribution of $H_{ijk\ell}(t)$ given $Y_{ijk\ell}(t) = y$ and $X_{ijk\ell}(t-1) = x$ is normal, with mean $(y\tau^2 x + w)/(1 + \tau^2 x)$ and variance $\tau^2/(1 + \tau^2 x)$. We will consider, the limiting posterior distribution obtained by letting $\tau^2 \to \infty$ (diffused prior). Under this condition

$$H_{ijk\ell}(t)\,|\,Y_{ijk\ell}(t), X_{ijk\ell}(t-1) \tag{5.15}$$

$$\sim N\left(Y_{ijk\ell}(t), \frac{1}{X_{ijk\ell}(t-1)}\right).$$

From (5.15) we obtain that the predictive distribution of $Y_{ijk\ell}(t+1)$, given $X_{ijk\ell}(t)$ and $X_{ijk\ell}(t-1)$ is

$$Y_{ijk\ell}(t+1)\,|\,X_{ijk\ell}(t), X_{ijk\ell}(t-1) \sim \tag{5.16}$$

$$N\left(Y_{ijk\ell}(t), \frac{1}{X_{ijk\ell}(t)} + \frac{1}{X_{ijk\ell}(t-1)}\right).$$

Hence, the <u>Bayes prediction limits</u> for $Y_{ijk\ell}(t+1)$, given $X_{ijk\ell}(t)$ and $X_{ijk\ell}(t-1)$ are

$$Y_{ijk\ell}(t) \pm z_{1-\alpha/2}\left\{\frac{1}{X_{ijk\ell}(t)} + \frac{1}{X_{ijk\ell}(t-1)}\right\}^{1/2} . \tag{5.17}$$

Notice that the interval given by (5.17) is always contained in the tolerance interval (5.9). If $w^{(U)}_{ijk\ell}$ and $w^{(L)}_{ijk\ell}(t)$ denote the upper and lower Bayes prediction limits in (5.17) then the Bayes prediction limits for $X_{ijk\ell}(t+1)$ are respectively

$$\zeta^{(U)}_{ijk\ell}(t) = (\sin(w^{(U)}_{ijk\ell}/2.))^2(X_{ijk\ell}(t) + 1) - .5 \tag{5.18}$$

and

$$\zeta^{(L)}_{ijk\ell}(t) = (\sin(w^{(L)}_{ijk\ell}/2.))^2(X_{ijk\ell}(t) + 1) - .5 \tag{5.19}$$

In the following table we provide numerical comparisons of the CMLE, tolerance and Bayes prediction limits for the size of cohorts in Phase II ($t > 2$) on July 1, 1972. We provide also the actual size of the cohorts.

TABLE 2

PREDICTION LIMITS FOR THE SIZE OF COHORTS ON JULY 1, 1972 (α = .05).

i	j	k	m	CMLE		Tolerance		Bayes		Actual
1	1	1	1	1401.	1434.	1382.	1446.	1393.	1439.	1412.
1	1	2	1	507.	520.	496.	523.	502.	521.	499.
1	2	1	1	3272.	3303.	3253.	3315.	3264.	3307.	3280.
1	2	2	1	676.	693.	665.	697.	671.	694.	674.
2	1	1	1	935.	979.	912.	997.	926.	986.	1009.
2	1	1	2	876.	916.	855.	932.	867.	922.	910.
2	1	1	3	1133.	1165.	1115.	1177.	1125.	1170.	1155.
2	1	2	1	113.	130.	103.	136.	109.	132.	135.
2	1	2	2	98.	111.	89.	114.	94.	112.	104.
2	1	[illegible]	3	158.	169.	149.	171.	154.	170.	163.
2	2	1	1	1973.	2009.	1952.	2023.	1964.	2014.	2019.
2	2	1	2	2277.	2309.	2257.	2322.	2269.	2314.	2301.
2	2	1	3	1112.	1133.	1099.	1140.	1106.	1136.	1121.
2	2	2	1	246.	260.	235.	264.	241.	261.	256.
2	2	2	2	223.	236.	214.	239.	219.	237.	230.
2	2	2	3	171.	181.	163.	183.	167.	182.	177.
3	1	1	1	1467.	1518.	1440.	1540.	1456.	1527.	1570.
3	1	1	2	1692.	1749.	1663.	1773.	1630.	1758.	1781.
3	1	1	3	2261.	2325.	2223.	2352.	2247.	2335.	2347.
3	1	1	4	2554.	2616.	2522.	2643.	2541.	2627.	2633.
3	1	1	5	2952.	3001.	2924.	3022.	2940.	3010.	2979.
3	1	2	1	147.	166.	136.	173.	142.	169.	165.
3	1	2	2	167.	187.	156.	194.	162.	190.	190.
3	1	2	3	266.	291.	252.	300.	260.	294.	292.
3	1	2	4	279.	300.	266.	207.	274.	302.	294.
3	1	2	5	391.	404.	380.	403.	386.	405.	386.
3	2	1	1	3333.	3379.	3307.	3399.	3322.	3387.	3403.
3	2	1	2	4623.	4676.	4594.	4699.	4611.	4685.	4675.
3	2	1	3	3380.	3427.	3354.	3447.	3369.	3435.	3416.
3	2	1	4	4115.	4159.	4090.	4177.	4104.	4166.	4108.
3	2	1	5	3207.	3240.	3187.	3252.	3199.	3245.	3207.
3	2	2	1	331.	348.	319.	353.	325.	350.	345.
3	2	2	2	431.	452.	417.	459.	425.	455.	451.
3	2	2	3	411.	432.	397.	440.	405.	435.	428.
3	2	2	4	466.	488.	453.	496.	460.	491.	478.
3	2	2	5	526.	540.	516.	544.	522.	541.	526.
Sums				48170	49249	47537	49656	47902	49404	49130

The m values corresponding to the various i values of Table 2 index the entry periods, in order of presentation in Table 1, excluding the July 71 entry period. Accordingly, $i = 1$, $m = 1$ corresponds to 1-71 entry period; for $i = 2$, $m = 1$ corresponds to 1-70 entry period, etc. As seen in Table 2 the three different systems of prediction limits provide (on retrospective basis) good forecasting. The CMLE intervals are the shortest and the tolerance intervals are the largest. As expected, the CMLE intervals might be however too short and the actual coverage probability could be too small. Indeed, in Table 2 the actual cohort sizes fall outside the CMLE intervals in 15 out of 36 cases, while only in four cases they are not covered by the tolerance intervals. When we sum the lower and the upper limits over all cohorts we obtain more conservative prediction intervals for the total force in Phase II. We see in Table 2 that all the three types of prediction intervals (CMLE, tolerance and Bayes) yield total force intervals which cover the actual value.

6. Estimating the Retention Rate After Six Months of Service

Cohorts which at the time of forecasting have been only six months in service should be considered separately, since generally the assumption $\theta_{ijk\ell}(t) = \theta_{ijk\ell}(t+1)$ is invalid for $t = 1$. Indeed, these cohorts just start their Phase II of service. We can, however, estimate in these cases the retention probability $\theta_{ijk\ell}(2)$ on the basis of the retention rates of previous cohorts during the period $t = 2$ (from six till twelve months). We propose the usage of the ratio estimator

$$\hat{\theta}_{ijk\ell}(2) = \frac{\sum_{v=1}^{\ell-1} X_{ijkv}(2)}{\sum_{v=1}^{\ell-1} X_{ijkv}(1)} \tag{6.1}$$

for every $\ell \geq 2$. Due to the independence of the random variables $X_{ijk\ell}(1)$ and $X_{ijk\ell}(2)$ as ℓ varies, the conditional distribution of $\sum_{v=1}^{\ell-1} X_{ijkv}(2)$, given $\{X_{ijk1}(1),\ldots,X_{ijk,\ell-1}(1)\}$, is the binomial $B\left(\sum_{v=1}^{\ell-1} X_{ijkv}(1), \theta_{ijk\ell}(2)\right)$. We assume here that $\theta_{ijk\ell}(2)$ is the same for all entry periods ($\ell = 1,2,\ldots$). Accordingly we can obtain CMLE, tolerance and Bayes prediction limits for $Y_{ijk\ell}(2)$, by employing the methods of Section 5 with the ratio estimator (6.1) rather than the estimator (5.1). The CMLE prediction limits are obtained from (5.2) by substituting (6.1) for $\hat{\theta}_{ijk\ell}(t)$. To obtain the tolerance limits for $X_{ijk\ell}(2)$ we define

$$Y^*_{ijk\ell}(2) = 2\sin^{-1}\sqrt{\hat{\theta}_{ijk\ell}(2)} \tag{6.2}$$

$(1-\alpha)$-confidence intervals for $H_{ijk\ell}(2)$ are then given, for large samples, by

$$Y^*_{ijk\ell}(2) \pm z_{1-\alpha/2} \Big/ \sqrt{\sum_{v=1}^{\ell-1} X_{ijkv}(1)} \tag{6.3}$$

Tolerance limits for $Y_{ijk\ell}(2)$, which is the arcsin transform of $(X_{ijk\ell}(2) + .5)/(X_{ijk\ell}(1) + 1.)$, are then

$$Y^*_{ijk\ell}(2) + z_{1-\alpha/2}\left\{\frac{1}{\sqrt{\sum_{v=1}^{\ell-1} X_{ijkv}(1)}} + \frac{1}{\sqrt{X_{ijk\ell}(2)}}\right\} \tag{6.4}$$

Substituting in (5.10)-(5.11) the two limits obtained from (6.4) we obtain tolerance limits for $X_{ijk\ell}(2)$. In a similar manner we can obtain that the Bayes prediction limits are

$$Y^*_{ijk\ell}(2) \pm z_{1-\alpha/2} \left\{ \frac{1}{\sum_{v=1}^{\ell-1} X_{ijkv}(1)} + \frac{1}{X_{ijk\ell}(1)} \right\}^{1/2} \tag{6.5}$$

Tolerance and Bayes prediction limits for $X_{ijk\ell}(2)$ are obtained by substituting the upper and lower limits of (6.4) and (6.5), respectively, in (5.10) and (5.11), or in (5.18) and (5.19). We try now this approach on the real Marine Corps data of first enlistees. In Table 3 we provide the number of first enlistees according to their cohorts which remained in service at epochs $t = 1$ and $t = 2$. We apply then the above formulae to forecast the numbers at $t = 2$ on the basis of $t = 1$.

TABLE 3

NUMBER OF ENLISTEES REMAINING IN SERVICE
AFTER 6 AND 12 MONTHS

Entry Period ℓ	Contract Length	LHS - W		LHS - N		AHS - W		AHS - N	
(1969)		6m	12m	6m	12m	6m	12m	6m	12m
1	2 Yrs.	6,057	5,631	1,647	1,544	12,116	11,648	2,127	2,017
2		3,421	3,146	1,175	1,080	10,756	10,385	1,722	1,637
1	3 Yrs.	1,767	1,627	204	194	2,374	2,312	211	196
2		2,376	2,166	298	270	3,603	3,501	456	433
1	4 Yrs.	3,145	2,853	441	408	4,217	4,081	471	456
2		3,197	2,864	371	334	5,547	5,371	586	554
(1970)									
3	2 Yrs.	1,440	1,373	496	474	3,915	3,823	869	843
4		1,562	1,493	421	400	4,269	4,169	790	769
3	3 Yrs.	1,466	1,372	195	187	2,496	2,228	302	292
4		1,208	1,143	134	131	2,496	2,437	257	251
3	4 Yrs.	3,370	3,165	432	401	3,934	3,835	532	512
4		3,333	3,173	371	356	4,527	4,401	563	547
(1971)									
5	2 Yrs.	1,564	1,489	537	525	3,416	3,351	720	702
6		1,809	1,717	744	716	4,553	4,451	1,235	1,195
5	3 Yrs.	1,293	1,219	181	172	1,180	1,151	190	183
6		919	866	143	133	1,034	997	233	227
5	4 Yrs.	3,325	3,146	423	410	3,366	3,294	539	546
6		3,359	3,174	402	383	4,445	4,281	632	607
(1972)									
7	2 Yrs.	2,710	2,585	1,223	1,177	3,517	3,413	1,311	1,273
8		1,659	1,576	729	688	3,149	3,075	1,052	1,014
7	3 Yrs.	2,067	1,964	434	417	1,329	1,299	324	307
8		1,766	1,669	465	441	1,568	1,528	500	477
7	4 Yrs.	4,945	4,696	940	893	3,731	3,620	684	655
8		4,680	4,407	1,577	1,514	4,461	4,305	1,393	1,328

Table 4

Prediction Limits for the Size of Cohorts at the Beginning of Phase II. (α = .05).

i	j	k	ℓ	CLE		Tolerance		Bayes		Actual
1	1	1	2	3151.	3210.	3127.	3229.	3142.	3216.	3146.
1	1	1	3	1314.	1353.	1305.	1359.	1312.	1353.	1373.
1	1	1	4	1432.	1472.	1423.	1478.	1430.	1472.	1493.
1	1	1	5	1440.	1478.	1431.	1484.	1438.	1479.	1489.
1	1	1	6	1671.	1712.	1662.	1718.	1669.	1712.	1717.
1	1	1	7	2514.	2563.	2502.	2572.	2511.	2564.	2585.
1	1	1	8	1539.	1577.	1532.	1581.	1537.	1577.	1576.
1	1	2	2	1085.	1118.	1069.	1129.	1079.	1121.	1080.
1	1	2	3	450.	472.	444.	475.	448.	472.	474.
1	1	2	4	383.	403.	378.	405.	382.	403.	400.
1	1	2	5	491.	514.	485.	516.	490.	513.	525.
1	1	2	6	687.	713.	680.	716.	685.	713.	716.
1	1	2	7	1139.	1170.	1129.	1176.	1136.	1171.	1177.
1	1	2	8	679.	703.	673.	705.	677.	702.	688.
1	2	1	2	10301.	10380.	10261.	10413.	10285.	10393.	10385.
1	2	1	3	3748.	3794.	3737.	3802.	3745.	3795.	3823.
1	2	1	4	4097.	4144.	4086.	4152.	4094.	4145.	4169.
1	2	1	5	3282.	3323.	3274.	3328.	3280.	3323.	3351.
1	2	1	6	4385.	4431.	4375.	4438.	4383.	4432.	4451.
1	2	1	7	3389.	3429.	3381.	3434.	3387.	3429.	3413.
1	2	1	8	3034.	3072.	3027.	3075.	3032.	3071.	3075.
1	2	2	2	1615.	1651.	1596.	1664.	1607.	1656.	1637.
1	2	2	3	812.	838.	804.	842.	810.	838	843.
1	2	2	4	741.	765.	735.	768.	739.	765.	769.
1	2	2	5	678.	699.	672.	701.	676.	699.	702.
1	2	2	6	1170.	1197.	1162.	1201.	1167.	1197.	1195.
1	2	2	7	1244.	1272.	1237.	1276.	1242.	1272.	1273.
1	2	2	8	999.	1024.	993.	1026.	998.	1023.	1014.
2	1	1	2	2162.	2214.	2128.	2240.	2146.	2225.	2166.
2	1	1	3	1321.	1363.	1307.	1373.	1317.	1365.	1372.
2	1	1	4	1094.	1131.	1084.	1138.	1091.	1132.	1143.
2	1	1	5	1178.	1215.	1168.	1221.	1175.	1216.	1219.
2	1	1	6	838.	868.	831.	872.	836.	868.	866.
2	1	1	7	1899.	1944.	1886.	1953.	1895.	1946.	1964.
2	1	1	8	1628.	1669.	1618.	1675.	1625.	1670.	1669.
2	1	2	2	276.	291.	263.	295.	270.	293.	270.
2	1	2	3	173.	187.	166.	190.	171.	188.	187.
2	1	2	4	120.	131.	115.	131.	118.	130.	131.
2	1	2	5	164.	177.	159.	178.	162.	176.	172.
2	1	2	6	129.	140.	126.	141.	128.	140.	133.
2	1	2	7	399.	418.	391.	422.	396.	419.	417.
2	1	2	8	431.	450.	423.	453.	428.	450.	441.
2	2	1	2	3490.	3528.	3463.	3546.	3477.	3536.	3501.
2	2	1	3	2412.	2444.	2399.	2451.	2407.	2445.	2228.
2	2	1	4	2347.	2390.	2333.	2400.	2343.	2392.	2437.
2	2	1	5	1113.	1141.	1107.	1144.	1112.	1141.	1151.

Table 4 (continued)

i	j	k	ℓ	CLE		Tolerance		Bayes		Actual
2	2	1	6	977.	1002.	972.	1005.	976.	1002.	997.
2	2	1	7	1258.	1287.	1252.	1290.	1257.	1287.	1299.
2	2	1	8	1489.	1520.	1483.	1523.	1488.	1520.	1528.
2	2	2	2	413.	434.	392.	445.	402.	440.	433.
2	2	2	3	277.	293.	269.	296.	274.	293.	292.
2	2	2	4	237.	251.	232.	253.	236.	251.	251.
2	2	2	5	176.	187.	172.	188.	175.	187.	183.
2	2	2	6	217.	229.	213.	230.	215.	228.	227.
2	2	2	7	304.	318.	299.	319.	302.	317.	307.
2	2	2	8	470.	488.	463.	490.	468.	488.	477.
3	1	1	2	2868.	2932.	2832.	2962.	2853.	2944.	2864.
3	1	1	3	3004.	3072.	2977.	3094.	2995.	3079.	3165.
3	1	1	4	3017.	3080.	2996.	3096.	3010.	3084.	3173.
3	1	1	5	3043.	3103.	3026.	3116.	3038.	3105.	3146.
3	1	1	6	3090.	3148.	3075.	3160.	3086.	3150.	3174.
3	1	1	7	4571.	4640.	4551.	4656.	4566.	4644.	4696.
3	1	1	8	4343.	4409.	4327.	4422.	4339.	4411.	4407.
3	1	2	2	333.	353.	321.	359.	328.	355.	334.
3	1	2	3	383.	406.	373.	412.	379.	408.	401.
3	1	2	4	331.	351.	323.	355.	328.	352.	356.
3	1	2	5	382.	403.	375.	406.	380.	403.	410.
3	1	2	6	367.	386.	361.	389.	365.	386.	383.
3	1	2	7	869.	897.	858.	904.	865.	899.	893.
3	1	2	8	1468.	1504.	1453.	1514.	1463.	1507.	1514.
3	2	1	2	5342.	5394.	5309.	5419.	5327.	5405.	5371.
3	2	1	3	3787.	3830.	3771.	3841.	3781.	3833.	3835.
3	2	1	4	4369.	4414.	4354.	4424.	4364.	4416.	4401.
3	2	1	5	3248.	3286.	3237.	3292.	3245.	3287.	3294.
3	2	1	6	4298.	4341.	4286.	4349.	4295.	4342	4281.
3	2	1	7	3600.	3641.	3591.	3647.	3597.	3641.	3620.
3	2	1	8	4306.	4351.	4296.	4358.	4304.	4351.	4305.
3	2	2	2	559.	576.	546.	581.	553.	578.	554.
3	2	2	3	499.	518.	490.	522.	496.	518.	512.
3	2	2	4	530.	549.	522.	552.	527.	549.	547.
3	2	2	5	509.	527.	503.	529.	507.	527.	546.
3	2	2	6	606.	622.	600.	624.	604.	622.	607.
3	2	2	7	654.	672.	649.	674.	652.	672.	655.
3	2	2	8	1335.	1361.	1325.	1366.	1332.	1362.	1328.

We see in Table 4 that the methods suggested in the present section provide prediction intervals which compare very well with the actual data.

REFERENCES

[1] Aitchison, J. and Dunsmore, I. R. (1975). Statistical Prediction Analysis Cambridge University Press, Cambridge.

[2] Haber, S. E. (1975). Factors influencing attrition in the Marine Corps. Technical Paper Serial T-306, Program in Logistics, The George Washington University.

[3] Johnson, N. L. and Kotz, S. (1969). Distributions in Statistics: Discrete Distributions. Houghton Mifflin Company, New York.

[4] Zacks, S. and Haber, S. E. (1975). A procedure for forecasting the size of a force subject to random withdrawals. Technical Paper Serial T-312, Program in Logistics, The George Washington University.

[5] Zacks, S. and Haber, S. E. (1975). Forecasting of Marine Corps force size as of December 31, 1974. Technical Memorandum Serial TM-80769, Program in Logistics, The George Washington University.

[6] Zacks, S. (1976). Forecasting the total size of the Marine Corps at the beginning of April, July and October 1976. Technical Memorandum, Series TM-65705, Program in Logistics, The George Washington University.

1. FORTRAN PROGRAM FOR THE DETERMINATION OF PREDICTION LIMITS; PHASE II COHORTS, t ≥ 2

```
100          DIMENSION NX(3,2,2,6,3)
110          Z=1.96
120          IA=3
130          JE=2
140          KR=2
150          MT=3
160          DO 1 I=1,IA
170          LA=2*I-1
180          DO 2 L=1,LA
190          READ(50,5) (((NX(I,J,K,L,M),M=1,MT),K=1,KR),J=1,JE)
200      5   FORMAT(12I4)
210      2   CONTINUE
220      1   CONTINUE
230          TFA=0.
240          TFL=0.
250          TFU=0.
260          TGL=0.
270          TGU=0.
280          THL=0.
290          THU=0.
300          DO 6 I=1,IA
310          DO 7 J=1,JE
320          DO 8 K=1,KR
325          LA=I*2-1
333          DO 9 L=1,LA
335          W=NX(I,J,K,L,1)
337          X=NX(I,J,K,L,2)
339          FA=NX(I,J,K,L,3)
340          TFA=TFA+FA
350          D=X*X/W
360          E=SQRT(ABS(D*(1.-X/W)))
370          FL=D-Z*E
380          FU=D+Z*E
390          TFL=TFL+FL
400          TFU=TFU+FU
410          P=(X+.5)/(W+1.)
420          Q=1.-P
430          R=SQRT(P/Q)
440          Y=2.*ATAN(R)
450          SE=SQRT(1./X)+SQRT(1./W)
460          CL=Y-Z*SE
470          CU=Y+Z*SE
480          GL=((SIN(CL/2.))**2)*(X+1.)-.5
485          GU=((SIN(CU/2.))**2)*(X+1.)-.5
490          TGL=TGL+GL
495          TGU=TGU+GU
500          SH=SQRT(1./X+1./W)
```

(cont'd)

```
      BL=Y-Z*SH
      BU=Y+Z*SH
      HL=((SIN(BL/2.))**2)*(X+1.)-.5
      HU=((SIN(BU/2.))**2)*(X+1.)-.5
      THL=THL+HL
      THU=THU+HU
      WRITE(66,10) I,J,K,L,FL,FU,GL,GU,HL,HU,FA
   10 FORMAT(4I3,7F7.0)
    9 CONTINUE
    8 CONTINUE
    7 CONTINUE
    6 CONTINUE
      WRITE(66,11) TFL,TFU,TGL,TGU,THL,THU,TFA
   11 FORMAT (12 X, 7F7.0)
      END
```

2. FORTRAN PROGRAM FOR THE DETERMINATION OF PREDICTION LIMITS; PHASE II COHORTS, t = 1.

```
      DIMENSION NX(3,2,2,8),NY(3,2,2,8)
      Z=1.96
      IA=3
      JE=2
      KR=2
      LA=8
     DO 1 M=1,4
     MM=M-1
     L=1+MM*2
     LL=L+1
      DO 2 I=1,IA
     READ(50,3) ((NX(I,J,K,L),NY(I,J,K,L),K=1,KR),J=1,JE)
      READ(50,3) ((NX(I,J,K,LL),NY(I,J,K,LL),K=1,KR),J=1,JE)
    3 FORMAT(8I5)
    2 CONTINUE
    1 CONTINUE
      DO 4 I=1,IA
      DO 5 J=1,JE
      DO 6 K=1,KR
      SX=0.
      SY=0.
      DO 7 L=2,LA
      LL=L-1
      X=NX(I,J,K,LL)
      SX=SX+X
      FX=NX(I,J,K,L)
      Y=NY(I,J,K,LL)
      SY=SY+Y
      FA=NY(I,J,K,L)
      TET=SY/SX
      D=FX*TET
      V=D*(1.-TET)
      SE=SQRT(V)
      FL=D-Z*SE
      FU=D+Z*SE
      R=TET/(1.-TET)
      SR=SQRT(R)
      W=2.*ATAN(SR)
      SW=SQRT(1./SX)+SQRT(1./FX)
      CL=W-Z*SW
      CU=W+Z*SW
      GL=((SIN(CL/2.))**2)*FX
      GU=((SIN(CU/2.))**2)*FX
      SV=SQRT(1./SX+1./FX)
      BL=W-Z*SV
      BU=W+Z*SV
      HL=((SIN(BL/2.))**2)*FX
      HU=((SIN(BU/2.))**2)*FX
      WRITE(66,10) I,J,K,L,FL,FU,GL,GU,HL,HU,FA
   10 FORMAT(4I3,7F7.0)
    7 CONTINUE
    6 CONTINUE
    5 CONTINUE
    4 CONTINUE
      END
```

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: (14) TR-27

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): (6) ADAPTIVE FORECASTING OF THE SIZE OF A FORCE ~~SUBST~~ SUBJECT TO RANDOM WITHDRAWALS.

5. TYPE OF REPORT & PERIOD COVERED: (9) Technical rept.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): (10) S. ZACKS AND S. E. HABER

8. CONTRACT OR GRANT NUMBER(s): (15) [redacted]N00014-75-C-0529
Project NR 042-276

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
CASE WESTERN RESERVE UNIVERSITY
DEPARTMENT OF MATHEMATICS AND STATISTICS

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (11) 15 Oct 76

11. CONTROLLING OFFICE NAME AND ADDRESS:
OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA

12. REPORT DATE: October 15, 1976

13. NUMBER OF PAGES: 28 (12) 30p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF DOCUMENT UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The study develops a new adaptive procedure for the forecasting of the total ~~siz~~ size of a military organization (the Marine Corps), subjected to random withdrawals of first enlistees and of career force personnel, with unknown attrition rates. The attrition rates of various subpopulations are not the same. The total population of first enlistees are classified to homogeneous subpopulations according to the following factors: length of contract, period of entrance to service, education and race. Based on cohorts having the same

→ cont. next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

403 901

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

bpz

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

cont. levels of these factors prediction intervals (with 95% confidence level) for the future size of these cohorts are estimated. Three methods of estimating prediction intervals are compared: conditional maximum likelihood, tolerance intervals and Bayes prediction intervals. Analysis of actual Marine Corps data illustrates the effectiveness of the forecasting procedures.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)